Nonlinear Precoding for Phase-Quantized Constant-Envelope Massive MU-MIMO-OFDM

Nonlinear Precoding for Phase-Quantized Constant-Envelope Massive MU-MIMO-OFDM

Sven Jacobsson[1,2], Oscar Castañeda[3], Charles Jeon[3], Giuseppe Durisi[1], and Christoph Studer[3]
[1]Chalmers University of Technology, Gothenburg, Sweden
[2]Ericsson Research, Gothenburg, Sweden
[3]Cornell University, Ithaca, NY, USA

***Abstract*—We propose a nonlinear phase-quantized constant-envelope precoding algorithm for the massive multi-user (MU) multiple-input multiple-output (MIMO) downlink. Specifically, we adapt the squared-infinity norm Douglas-Rachford splitting (SQUID) precoder to systems that use oversampling digital-to-analog converters (DACs) at the base station (BS) and orthogonal frequency-division multiplexing (OFDM) to communicate over frequency-selective channels. We demonstrate that the proposed SQUID-OFDM precoder is able to generate transmit signals that are constrained to constant envelope, which enables the use of power-efficient analog radio-frequency circuitry at the BS. By quantizing the phase of the resulting constant-envelope signal, we obtain a finite-cardinality transmit signal that can be synthesized by low-resolution (e.g., 1-bit) DACs. We use error-rate simulations to demonstrate the superiority of SQUID-OFDM over linear-quantized precoders for massive MU-MIMO-OFDM systems.**

I. Introduction

Massive multi-user (MU) multiple-input multiple-output (MIMO) equips the base station (BS) with a large number of antenna elements and serves tens of user equipments (UEs) simultaneously and in the same frequency band [1], [2]. While massive MU-MIMO is expected to be a key technology component of fifth-generation (5G) wireless networks, scaling traditional radio frequency (RF) front-end architectures to BSs with hundreds of antenna elements leads to a prohibitive growth in circuit power consumption, system costs, and hardware complexity. Hence, a successful deployment of massive MU-MIMO requires inexpensive, power-efficient, and low-complexity hardware components, which, in turn, will limit the capacity of the system due to signal-quality degradation.

A. Constant-Envelope and Phase-Quantized Precoding

In the massive MU-MIMO downlink (the BS transmits data to the UEs), precoding must be used to reduce MU interference. Unfortunately, precoding typically generates time-domain signals with high peak-to-average power ratio (PAR) [3]; this fact is further aggravated in systems that use orthogonal frequency-division multiplexing (OFDM) to facilitate communication over wideband frequency-selective channels [4]. For such high-PAR waveforms, one has to operate the power amplifiers (PAs) in the linear regime to prevent significant signal-quality degradation. This results in high PA power consumption [4].

To mitigate the high-PAR issue, a constant-envelope precoder for the massive MU-MIMO-OFDM case was proposed in [5]; its design ensures that the precoded signal has equal amplitude on all antennas (and, hence, zero PAR). This precoder enables PAs to operate in the nonlinear regime, allowing for energy-efficient analog circuitry. Recently, the authors of [6] designed a precoder for the frequency-flat case that outputs a constant-envelope signal constrained to only eight phases. This precoder requires the digital-to-analog converters (DACs) at the BS to generate only eight phase outputs, which enables the use of power-efficient converter architectures, and reduces the interconnect data rates between the baseband-processing unit and the radio unit at the BS.

B. 1-Bit Precoding

Motivated by potential power savings and reduced interconnect data rates, the use of 1-bit DACs in the massive MU-MIMO downlink has recently attracted significant attention. Specifically, so-called *linear-quantized* precoders (i.e., linear precoding followed by quantization) have been recently proposed for precoding in massive MU-MIMO-OFDM systems that use oversampling DACs [7], [8]. These precoders achieve low bit-error rates (BERs) and high sum-rate throughputs over frequency-selective channels with OFDM, despite the adverse impact of the 1-bit DACs. *Nonlinear* precoders, where the precoder depends on the instantaneous realizations of the information symbols, are known to significantly outperform linear-quantized precoders (see, e.g., [9]–[15]), but have, until recently, been analyzed exclusively for frequency-flat channels and single-carrier transmission.[1]

C. Contributions

We propose a nonlinear phase-quantized constant-envelope precoder for the massive MU-MIMO-OFDM downlink operating over frequency-selective channels. Our precoder builds upon the squared-infinity-norm Douglas-Rachford splitting (SQUID) algorithm put forward in [11, Sec. IV-B]. In contrast to previous works [9]–[15], which focus on the case of Nyquist-rate sampling 1-bit DACs and on the frequency-flat case, the proposed

The work of SJ and GD was supported in part by the Swedish Foundation for Strategic Research under grant ID14-0022, and by the Swedish Governmental Agency for Innovation Systems (VINNOVA) within the competence center ChaseOn. SJ's research visit at Cornell was sponsored in part by Cornell's College of Engineering. The work of OC, CJ, and CS was supported by Xilinx, Inc. and by the US National Science Foundation (NSF) under grants ECCS-1408006, CCF-1535897, CAREER CCF-1652065, and CNS-1717559.

[1]See, however, [16] for a recent result on the frequency-selective case with OFDM transmission.

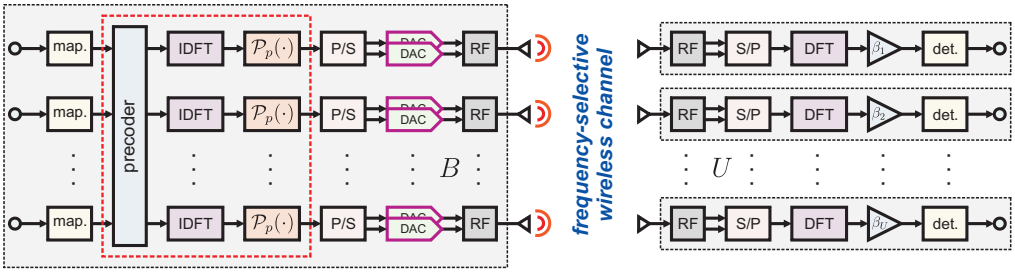


Fig. 1. Overview of the considered massive MU-MIMO-OFDM downlink system. Left: BS with B antennas performs precoding in the frequency domain, transforms the precoded vector into time domain, and maps its entries to the set of outcomes supported by the transcoder in the DACs. The dashed red box indicates the operations carried out by the nonlinear phase-quantized constant-envelope precoder. Right: U single-antenna UEs.

nonlinear precoder, which we shall refer to as SQUID-OFDM, is capable of supporting oversampling DACs and OFDM. We characterize the computational complexity of SQUID-OFDM and demonstrate its efficacy via numerical simulations.

D. Notation

Lowercase and uppercase boldface letters denote vectors and matrices, respectively. The $M \times N$ all-zeros matrix and the $M \times M$ identity matrix are denoted by $\mathbf{0}_{M\times N}$ and $\mathbf{I}_M$, respectively. The unitary $N \times N$ discrete Fourier transform (DFT) matrix is denoted by $\mathbf{F}_N$. The ℓ_∞-norm of $\mathbf{a} = [a_1, \ldots, a_M]^T$ is $\|\mathbf{a}\|_\infty = \max_{\ell=1,\ldots,M}|a_\ell|$; the $\ell_{\widetilde{\infty}}$-norm is $\|\mathbf{a}\|_{\widetilde{\infty}} = \max\left\{\|\Re\{\mathbf{a}\}\|_\infty, \|\Im\{\mathbf{a}\}\|_\infty\right\}$. We use $\|\mathbf{a}\|_2$ and $\|\mathbf{A}\|_F$ to denote the ℓ_2-norm of vector $\mathbf{a}$ and the Frobenius norm of matrix $\mathbf{A}$, respectively. If $\mathbf{A}$ is an $M \times N$ matrix, then $\mathrm{vec}(\mathbf{A})$ is an MN-dimensional vector obtained by column-wise vectorization of $\mathbf{A}$. The phase of $a \in \mathbb{C}$ is denoted by $\arg(a)$; the sign of $r \in \mathbb{R}$ is denoted by $\mathrm{sgn}(r) \in \{-1, +1\}$. The floor function $\lfloor r \rfloor$ produces the largest integer less than or equal to r. The complex-valued circularly symmetric Gaussian distribution with covariance matrix $\mathbf{K} \in \mathbb{C}^{M\times M}$ is denoted by $\mathcal{CN}(\mathbf{0}_{M\times 1}, \mathbf{K})$. The expected value of $\mathbf{A}$ is $\mathbb{E}[\mathbf{A}]$.

II. System Model

We consider a single-cell massive MU-MIMO-OFDM downlink system as illustrated in Fig. 1. The system operates over a wideband channel where OFDM is used to deal with the selectiveness in frequency of the channel. Let B denote the number of BS antennas and U the number of single-antenna UEs. At the BS, the frequency-domain information symbols are mapped to the antenna array by a precoder. At each BS antenna, the precoded signal is mapped to time domain through an inverse discrete Fourier transform (IDFT) before being passed to a pair of finite-resolution DACs, which generate the in-phase and quadrature components of the transmitted time-domain signal. For simplicity, we ignore other RF impairments and assume perfect synchronization between the BS and the UEs.

A. Channel Input-Output Relation

Under the above assumptions, the received signal $\mathbf{y}_n \in \mathbb{C}^U$ at the U UEs can be written as

$$\mathbf{y}_n = \sum_{\ell=0}^{L-1} \mathbf{H}_\ell \mathbf{x}_{n-\ell} + \mathbf{w}_n \tag{1}$$

at discrete time instants $n = 0, \ldots, N-1$. Here, $\mathbf{x}_n$ is the B-dimensional transmit signal at discrete time n and N is the number of samples per OFDM symbol (the size of the IDFT). The vector $\mathbf{w}_n \sim \mathcal{CN}(\mathbf{0}_{U\times 1}, N_0\mathbf{I}_U)$ denotes the i.i.d. additive white Gaussian noise (AWGN) at the UEs at discrete time n. Here, N_0 is the noise power and $SNR = 1/N_0$ defines the signal-to-noise ratio (SNR). The matrix $\mathbf{H}_\ell \in \mathbb{C}^{U\times B}$ is the ℓth tap of the frequency-selective channel ($\ell = 0, \ldots, L-1$). We assume that the realizations of $\{\mathbf{H}_\ell\}_{\ell=0}^{L-1}$ remain constant for the duration of each OFDM symbol and that they are perfectly known to the BS. Let $\mathbf{X} = [\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}]^T$, $\mathbf{Y} = [\mathbf{y}_0, \ldots, \mathbf{y}_{N-1}]^T$, and $\mathbf{W} = [\mathbf{w}_0, \ldots, \mathbf{w}_{N-1}]^T$. Furthermore, we let $\widehat{\mathbf{X}} = \mathbf{X}\mathbf{F}_N^H$, $\widehat{\mathbf{Y}} = \mathbf{Y}\mathbf{F}_N^H$, $\widehat{\mathbf{W}} = \mathbf{W}\mathbf{F}_N^H$, and $\widehat{\mathbf{H}}_k = \sum_{\ell=0}^{L-1} \mathbf{H}_\ell e^{-jk\frac{2\pi}{N}\ell}$. A cyclic prefix of length $L-1$ is prepended to the transmit signal at the BS. After removing the cyclic prefix and after a DFT at the UEs, the received signal at the U UEs and on the kth subcarrier can be written as

$$\hat{\mathbf{y}}_k = \widehat{\mathbf{H}}_k \hat{\mathbf{x}}_k + \hat{\mathbf{w}}_k \tag{2}$$

for $k = 0, \ldots, N-1$. Here, $\hat{\mathbf{x}}_k$, $\hat{\mathbf{y}}_k$, and $\hat{\mathbf{w}}_k$ correspond to the kth column of $\widehat{\mathbf{X}}$, $\widehat{\mathbf{Y}}$, and $\widehat{\mathbf{W}}$, respectively.

B. Precoding, Quantization, and OFDM Parameters

We use the disjoint sets $\mathcal{I}$ and $\mathcal{G}$, where $|\mathcal{I}| + |\mathcal{G}| = N$, to denote the set of subcarriers associated with information symbols (occupied subcarriers) and zeros (guard subcarriers), respectively. We let $S = |\mathcal{I}|$ be the number of occupied subcarriers and define N/S as the oversampling ratio. Let $\mathbf{s}_k = [s_{1,k}, \ldots, s_{U,k}]^T$ denote the symbol vector associated with the kth subcarrier ($k = 0, \ldots, N-1$). We assume that $s_{u,k} \in \mathcal{O}$ for all $k \in \mathcal{I}$ and that $s_{u,k} = 0$ for all $k \in \mathcal{G}$. Here, $\mathcal{O}$ represents a quadrature amplitude modulation (QAM) constellation (e.g., 16-QAM),.

The precoder uses the available transmit-side channel-state information to map the symbols $\mathbf{S} = [\mathbf{s}_0, \ldots, \mathbf{s}_{N-1}] \in \mathcal{O}^{U\times N}$, to the transmitted signal $\mathbf{X}$, which must satisfy the average power constraint $\mathbb{E}_{\mathbf{S}}\left[\|\mathbf{X}\|_F^2\right] = S$. Due to the finite resolution of real-world DACs we require that $\mathbf{X} \in \mathcal{X}_p^{B\times N}$, where $\mathcal{X}_p$ is the set of values that are supported by the DACs. We shall assume that $\mathcal{X}_p$ is a constant-envelope alphabet and let $p > 0$ be the number of *phase bits*, so that 2^p is the number of possible phases of the signal transmitted at each antenna.

Furthermore, we let $|x|^2 = P_{\text{ant}}$, $x \in \mathcal{X}_p$. Here, $P_{\text{ant}} = S/(BN)$ is the per-antenna transmit power, which ensures that the average power constraint is satisfied. For $p < \infty$, the mth element ($m = 0, \ldots, 2^p - 1$) of the set $\mathcal{X}_p$ is hence given by $(P_{\text{ant}})^{1/2} e^{j(\pi + 2\pi m)/2^p}$. We let $\mathcal{X}_\infty = \{x \in \mathbb{C} : |x|^2 = P_{\text{ant}}\}$.

In this paper, we shall benchmark the performance of our nonlinear precoding algorithm, SQUID-OFDM, against the linear Wiener-filter (WF) precoder [17] given by

$$\mathbf{X}^{\text{WF}} = \mathcal{P}_p\Big(\widehat{\mathbf{Z}}^{\text{WF}} \mathbf{F}_N^H\Big), \tag{3}$$

where the kth column of $\widehat{\mathbf{Z}}^{\text{WF}}$ is given by

$$\hat{\mathbf{z}}_k^{\text{WF}} = \frac{1}{\beta^{\text{WF}}} \widehat{\mathbf{H}}_k^H \Big(\widehat{\mathbf{H}}_k \widehat{\mathbf{H}}_k^H + UN_0 \mathbf{I}_U\Big)^{-1} \mathbf{s}_k \tag{4}$$

for $k \in \mathcal{I}$, and by $\hat{\mathbf{z}}_k^{\text{WF}} = \mathbf{0}_{B\times 1}$ for $k \in \mathcal{G}$. Here, the constant $\beta^{\text{WF}} \in \mathbb{R}^+$ ensures that $\mathbb{E}_{\mathbf{S}}[\|\widehat{\mathbf{Z}}^{\text{WF}} \mathbf{F}_N^H\|_F^2] = S$. Prior to transmission, the time-domain precoded signal $\widehat{\mathbf{Z}}^{\text{WF}} \mathbf{F}_N^H$ in (3) is quantized by the function $\mathcal{P}_p(\cdot) : \mathbb{C}^{B\times N} \to \mathcal{X}_p^{B\times N}$, which is applied entry-wise to the matrix $\mathbf{X}^{\text{WF}}$, so that the transmitted signal matches the transcoder in the DACs. Specifically,

$$\mathcal{P}_p(z) = \begin{cases} \sqrt{P_{\text{ant}}} e^{j\frac{2\pi}{2^p}\left(\left\lfloor \frac{2^p \arg(z)}{2\pi} \right\rfloor + \frac{1}{2}\right)}, & p < \infty \\ \sqrt{P_{\text{ant}}} e^{j \arg(z)}, & p = \infty. \end{cases} \tag{5}$$

Note that for the 2-phase-bit case ($p = 2$), we retrieve from (5) the 1-bit-DAC setup studied in [8]. There, the in-phase and quadrature components of the per-antenna transmitted signal are generated independently by a pair of 1-bit-DAC and

$$\mathcal{P}_2(z) = \sqrt{\frac{P_{\text{ant}}}{2}} (\operatorname{sgn}(\Re\{z\}) + j \operatorname{sgn}(\Im\{z\})). \tag{6}$$

The 1-phase-bit case ($p = 1$), on the other hand, corresponds to the case when there is only *a single* 1-bit DAC per antenna, i.e., the transmitted signal has no in-phase component.

III. Nonlinear Constant-Envelope Precoding

As in [11], [12], we focus on a nonlinear precoding strategy that minimizes the mean square error (MSE) at the UEs. Let $MSE_{u,k} = \mathbb{E}_{w_{u,k}}\big[|s_{u,k} - \beta \hat{y}_{u,k}|^2\big]$ denote the MSE for the uth UE and on the kth subcarrier. Here, $\hat{y}_{u,k}$ is the uth element of $\hat{\mathbf{y}}_k$ and $\beta \in \mathbb{R}^+$ is a constant that takes into account the channel gain. With these definitions, we write the sum-MSE over the U UEs and over the S occupied subcarriers as

$$\sum_{u=1}^{U} \sum_{k\in\mathcal{I}} MSE_{u,k} = \sum_{k\in\mathcal{I}} \mathbb{E}_{\hat{\mathbf{w}}_k}\big[\|\mathbf{s}_k - \beta \hat{\mathbf{y}}_k\|_2^2\big] \tag{7}$$

$$= \sum_{k\in\mathcal{I}} \|\mathbf{s}_k - \beta \widehat{\mathbf{H}}_k \hat{\mathbf{x}}_k\|_2^2 + \beta^2 USN_0. \tag{8}$$

Recall that $\hat{\mathbf{x}}_k$ is the kth column of $\widehat{\mathbf{X}}$. We can now define the sum-MSE-optimal precoding problem (PP) as follows:

$$\text{(PP)} \quad \begin{cases} \underset{\mathbf{X}\in\mathcal{X}_p^{B\times N},\, \beta\in\mathbb{R}^+}{\text{minimize}} & \sum\limits_{k\in\mathcal{I}} \|\mathbf{s}_k - \beta \widehat{\mathbf{H}}_k \hat{\mathbf{x}}_k\|_2^2 + \beta^2 USN_0 \\ \text{subject to} & \mathbf{X} = \widehat{\mathbf{X}} \mathbf{F}_N^H. \end{cases} \tag{9}$$

For constant-envelope signals that adhere to the average power constraint, it holds that $\|\text{vec}(\mathbf{X})\|_\infty^2 = P_{\text{ant}}$, and the problem (PP) can equivalently be written as

$$\begin{array}{ll} \underset{\mathbf{X}\in\mathcal{X}_p^{B\times N},\, \beta\in\mathbb{R}^+}{\text{minimize}} & \sum\limits_{k\in\mathcal{I}} \|\mathbf{s}_k - \beta \widehat{\mathbf{H}}_k \hat{\mathbf{x}}_k\|_2^2 + \beta^2 \gamma \|\text{vec}(\mathbf{X})\|_\infty^2 \\ \text{subject to} & \mathbf{X} = \widehat{\mathbf{X}} \mathbf{F}_N^H, \end{array} \tag{10}$$

where $\gamma = BUNN_0$. Proceeding analogously to [11, Sec. IV-B], by setting $\widehat{\mathbf{B}} = \beta \widehat{\mathbf{X}}$ and by dropping the nonconvex constraint $\mathbf{X} \in \mathcal{X}_p^{B\times N}$, we obtain the following convex relaxation of the problem in (10), which we denote by (Pℓ_∞^2):

$$(\text{P}\ell_\infty^2) \quad \underset{\widehat{\mathbf{B}}\in\mathbb{C}^{B\times N}}{\text{minimize}} \sum_{k\in\mathcal{I}} \|\mathbf{s}_k - \widehat{\mathbf{H}}_k \hat{\mathbf{b}}_k\|_2^2 + \gamma \|\text{vec}\big(\widehat{\mathbf{B}} \mathbf{F}_N^H\big)\|_\infty^2. \tag{11}$$

Here, $\hat{\mathbf{b}}_k$ is the kth column of $\widehat{\mathbf{B}}$. Let $\widehat{\mathbf{B}}^{\text{P}\ell_\infty^2}$ and $\beta^{\text{P}\ell_\infty^2}$ denote the optimal solutions to the problem (Pℓ_∞^2). We obtain the desired matrix $\mathbf{X}^{\text{P}\ell_\infty^2}$ by converting $\widehat{\mathbf{B}}^{\text{P}\ell_\infty^2}$ to time-domain and by mapping the resulting matrix to the set $\mathcal{X}_p^{B\times N}$ using (5), i.e.,

$$\mathbf{X}^{\text{P}\ell_\infty^2} = \mathcal{P}_p\big(\widehat{\mathbf{B}}^{\text{P}\ell_\infty^2} \mathbf{F}_N^H\big). \tag{12}$$

In Section III-A, we will show that the problem (Pℓ_∞^2) can be solved efficiently. Note that for the 2-phase-bit case, it holds that $\|\text{vec}(\mathbf{X})\|_\infty^2 = 2\|\text{vec}(\mathbf{X})\|_{\widetilde{\infty}}^2$. In this case, it turns out that one achieves better performance by solving instead the following optimization problem, which we denote by (P$\ell_{\widetilde{\infty}}^2$):

$$(\text{P}\ell_{\widetilde{\infty}}^2) \quad \underset{\widehat{\mathbf{B}}\in\mathbb{C}^{B\times N}}{\text{minimize}} \sum_{k\in\mathcal{I}} \|\mathbf{s}_k - \widehat{\mathbf{H}}_k \hat{\mathbf{b}}_k\|_2^2 + 2\gamma \|\text{vec}\big(\widehat{\mathbf{B}} \mathbf{F}_N^H\big)\|_{\widetilde{\infty}}^2. \tag{13}$$

We shall discuss the implications of this slight modification of the precoding problem in the next section.

A. SQUID-OFDM Precoding

Douglas-Rachford splitting [18] is an efficient iterative scheme to solve convex optimization problems of the form

$$\underset{\widehat{\mathbf{B}}\in\mathbb{C}^{B\times N}}{\text{minimize}} \quad f\big(\widehat{\mathbf{B}}\big) + g\big(\widehat{\mathbf{B}}\big), \tag{14}$$

where $f(\cdot)$ and $g(\cdot)$ are closed convex functions, which have *proximal operators* [19] defined as follows:

$$\text{prox}_f(\mathbf{V}) = \underset{\widehat{\mathbf{B}}\in\mathbb{C}^{B\times N}}{\arg\min}\, f(\widehat{\mathbf{B}}) + \tfrac{1}{2}\|\widehat{\mathbf{B}} - \mathbf{V}\|_F^2 \tag{15}$$

$$\text{prox}_g(\mathbf{V}) = \underset{\widehat{\mathbf{B}}\in\mathbb{C}^{B\times N}}{\arg\min}\, g(\widehat{\mathbf{B}}) + \tfrac{1}{2}\|\widehat{\mathbf{B}} - \mathbf{V}\|_F^2. \tag{16}$$

By starting at an arbitrary $\widehat{\mathbf{B}}^{(0)}$ and $\widehat{\mathbf{C}}^{(0)}$, Douglas-Rachford splitting solves problems of the form (14) exactly [20] by repeating for $t = 1, \ldots, T$, where T is the maximum number of iterations, the following iterative procedure:

$$\widehat{\mathbf{A}}^{(t)} = \text{prox}_f\Big(2\widehat{\mathbf{B}}^{(t-1)} - \widehat{\mathbf{C}}^{(t-1)}\Big) \tag{17}$$

$$\widehat{\mathbf{B}}^{(t)} = \text{prox}_g\Big(\widehat{\mathbf{C}}^{(t-1)} + \widehat{\mathbf{A}}^{(t)} - \widehat{\mathbf{B}}^{(t-1)}\Big) \tag{18}$$

$$\widehat{\mathbf{C}}^{(t)} = \widehat{\mathbf{C}}^{(t-1)} + \widehat{\mathbf{A}}^{(t)} - \widehat{\mathbf{B}}^{(t)}. \tag{19}$$

We now outline the SQUID-OFDM precoder, which builds upon the SQUID precoder proposed in [11, Sec. IV-B] and

performs Douglas-Rachford splitting to solve the problems (Pℓ_∞^2) and (P$\ell_{\widetilde{\infty}}^2$). Specifically, SQUID-OFDM extends SQUID to support OFDM, oversampling DACs, and arbitrary constant-envelope alphabets. Let $f(\widehat{\mathbf{B}}) = \sum_{k\in\mathcal{I}}\|\mathbf{s}_k - \widehat{\mathbf{H}}_k\hat{\mathbf{b}}_k\|_2^2$. For the problem (P$\ell_\infty^2$), $g(\widehat{\mathbf{B}}) = \gamma\|\mathrm{vec}(\widehat{\mathbf{B}}\mathbf{F}_N^H)\|_\infty^2$. For the problem (P$\ell_{\widetilde{\infty}}^2$), $g(\widehat{\mathbf{B}}) = 2\gamma\|\mathrm{vec}(\widehat{\mathbf{B}}\mathbf{F}_N^H)\|_{\widetilde{\infty}}^2$. In both cases, the proximal operator for $g(\cdot)$ in (16) can computed using [11, Alg. 1]. For the proximal operator of $f(\cdot)$, we note that the objective function in (15) is separable in the columns of $\widehat{\mathbf{B}}$ and that the kth column of $\widehat{\mathbf{A}}^{(t)}$ in (17) can be computed as

$$\hat{\mathbf{a}}_k^{(t)} = \Big(\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \tfrac{1}{2}\mathbf{I}_B\Big)^{-1}\Big(\widehat{\mathbf{H}}_k^H\mathbf{s}_k + \hat{\mathbf{b}}_k^{(t-1)} - \tfrac{1}{2}\hat{\mathbf{c}}_k^{(t-1)}\Big) \quad (20)$$

$$= \big(\mathbf{I}_B - \mathbf{Q}_k\widehat{\mathbf{H}}_k\big)\big(2\hat{\mathbf{b}}_k^{(t-1)} - \hat{\mathbf{c}}_k^{(t-1)}\big) + \mathbf{d}_k \quad (21)$$

for $k\in\mathcal{I}$, and $\mathbf{a}_k^{(t)} = 2\hat{\mathbf{b}}_k^{(t-1)} - \hat{\mathbf{c}}_k^{(t-1)}$ for $k\in\mathcal{G}$. Here, $\mathbf{a}_k^{(t)}$ is the kth column of $\widehat{\mathbf{A}}^{(t)}$ and $\mathbf{c}_k^{(t)}$ is the kth column of $\widehat{\mathbf{C}}^{(t)}$. To derive (21), we used the Woodbury matrix identity to reduce the dimension of the inverse and to speed up computations by precomputing, for $k\in\mathcal{I}$, the matrix $\mathbf{Q}_k\in\mathbb{C}^{B\times U}$ and the vector $\mathbf{d}_k\in\mathbb{C}^B$, which are defined as follows:

$$\mathbf{Q}_k = \widehat{\mathbf{H}}_k^H\Big(\widehat{\mathbf{H}}_k\widehat{\mathbf{H}}_k^H + \frac{1}{2}\mathbf{I}_U\Big)^{-1} \quad (22)$$

$$\mathbf{d}_k = 2\Big(\widehat{\mathbf{H}}_k^H\mathbf{s}_k - \mathbf{Q}_k\widehat{\mathbf{H}}_k\widehat{\mathbf{H}}_k^H\mathbf{s}_k\Big). \quad (23)$$

We can now solve the problems (Pℓ_∞^2) and (P$\ell_{\widetilde{\infty}}^2$) by using the iterative procedure outlined in Algorithm 1.

Algorithm 1 (SQUID-OFDM): Compute $\mathbf{Q}_k$ and $\mathbf{d}_k$ for $k\in\mathcal{I}$ using (22) and (23), respectively. Initialize $\widehat{\mathbf{B}}^{(0)} = \mathbf{0}_{B\times N}$, and $\widehat{\mathbf{C}}^{(0)} = \mathbf{0}_{B\times N}$. Then, at every iteration $t = 1, 2, \ldots, T$, compute the following three quantities:

$$\hat{\mathbf{a}}_k^{(t)} = \begin{cases}\mathbf{R}_k\Big(2\hat{\mathbf{b}}_k^{(t-1)} - \hat{\mathbf{c}}_k^{(t-1)}\Big) + \mathbf{d}_k, & k\in\mathcal{I}\\ 2\hat{\mathbf{b}}_k^{(t-1)} - \hat{\mathbf{c}}_k^{(t-1)}, & k\in\mathcal{G}\end{cases} \quad (24)$$

$$\widehat{\mathbf{B}}^{(t)} = \mathrm{prox}_g\Big(\widehat{\mathbf{C}}^{(t-1)} + \widehat{\mathbf{A}}^{(t)} - \widehat{\mathbf{B}}^{(t-1)}\Big)\mathbf{F}_N \quad (25)$$

$$\widehat{\mathbf{C}}^{(t)} = \widehat{\mathbf{C}}^{(t-1)} + \widehat{\mathbf{A}}^{(t)} - \widehat{\mathbf{B}}^{(t)}. \quad (26)$$

Here, $\mathbf{R}_k = \big(\mathbf{I}_B - \mathbf{Q}_k\widehat{\mathbf{H}}_k\big)$. In (25), $\mathrm{prox}_g(\cdot)$ is computed using [11, Alg. 1]. After the last iteration, obtain the transmitted signal $\mathbf{X}^{(T)}$ by quantizing $\widehat{\mathbf{B}}^{(T)}\mathbf{F}_N^H$ to the constant-envelope alphabet $\mathcal{X}_p^{B\times N}$ using (5).

Fig. 2 shows the time-domain SQUID-OFDM output $\widehat{\mathbf{B}}^{(20)}\mathbf{F}_N^H$ after $T = 20$ iterations before and after quantization.[2] Fig. 2b shows the SQUID-OFDM output for the problem (P$\ell_{\widetilde{\infty}}^2$) before and after 2-phase-bit quantization.[3] We see that the $\ell_{\widetilde{\infty}}$-norm constrains the SQUID-OFDM output to a box in the complex plane, which limits the error caused by the quantizer. Fig. 2c shows the output for the problem (Pℓ_∞^2) before and after 3-phase-bit quantization. In this case, the ℓ_∞-norm constrains the SQUID-OFDM output to a circle in the complex plane, which is suitable for quantization using three phase bits or more. Fig. 2a shows the SQUID-OFDM output for the problem (Pℓ_∞^2) before and after 1-phase-bit quantization. Here, we slightly modified the problem to force the real part of the output to the proximal operator $\mathrm{prox}_g(\cdot)$ to zero. This constrains the SQUID-OFDM output to a line in the complex plane, which is suitable for quantization using only one phase bit.

[2]The simulation parameters are given in Section IV-A.

[3]By setting $S = 1$ and $N = 1$ for the problem (P$\ell_{\widetilde{\infty}}^2$), Algorithm 1 reduces to the SQUID precoder for single-carrier transmission [11, Sec. IV-B].

TABLE I
COMPLEXITY COMPARISON BETWEEN WF AND SQUID-OFDM.

Precoder	Computational complexity
WF	$2S\big(\frac{1}{3}U^3 + BU^2 + 2U^2 - \frac{1}{3}U\big)$ $+4B(N\log_2 N - 3N + 4)$
SQUID-OFDM	$2S\big(\frac{5}{3}U^3 + 3BU^2 + \big(6B - \frac{2}{3}\big)U\big)$ $+4TB(2SU + 2N\log_2 N - 5N + 8)$

B. Computational Complexity

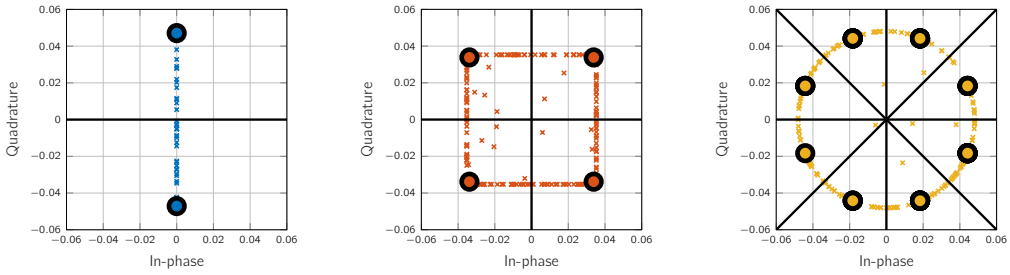
Table I shows the computational complexity characterized by the number of *real-valued multiplications* for SQUID-OFDM and WF precoding. In what follows, we assume that one complex-valued multiplication requires four real-valued multiplications.

1) WF Precoding: Computing (4) exactly for all $k\in\mathcal{I}$ using implicit Cholesky-based matrix inversion [21] requires $2S\big(\frac{1}{3}U^3 + BU^2 + 2U^2 - \frac{1}{3}U\big)$ real-valued multiplications. At each antenna element, the frequency-domain precoded vector is converted to time-domain via an IDFT. Computing these IDFTs require $4B(N\log_2 N - 3N + 4)$ real-valued multiplications if the IDFTs are computed using the split-radix Fast Fourier Transform (FFT) algorithm [22, Sec. 4.3]. By adding these two numbers, we obtain the complexity reported in Table I.

2) SQUID-OFDM: The preprocessing step of SQUID-OFDM involves computing (22) and (23) for $k\in\mathcal{I}$. Proceeding as in [23], we find that computing $\mathbf{Q}_k$ in (22) for $k\in\mathcal{I}$ requires $2S(\frac{5}{3}U^3 + 3BU^2 - \frac{2}{3}U)$ real-valued multiplications. Furthermore, computing $\mathbf{d}_k$ in (23) for $k\in\mathcal{I}$ requires an additional $12SBU$ real-valued multiplications. By adding these numbers, we find that the preprocessing complexity of SQUID-OFDM is $2S\big(\frac{5}{3}U^3 + 3BU^2 + \big(6B - \frac{2}{3}\big)U\big)$. Moving on to the per-iteration complexity. Computing efficiently the vectors $\mathbf{a}_k^{(t)}$ in (24) for $k\in\mathcal{I}$ requires $8SBU$ real-valued multiplications per iteration.[4] Furthermore, executing [11, Alg. 1] requires $4BN$ real-valued multiplications, which means that computing $\widehat{\mathbf{B}}^{(t)}$ in (25), if the split-radix FFT algorithm is used to compute the IDFT and DFT, requires an additional $4B(2N\log_2 N - 5N + 8)$ real-valued multiplications per iteration. Hence, the per-iteration complexity of SQUID-OFDM is $4B(2SU + 2N\log_2 N - 5N + 8)$.[5] Finally, by adding the

[4]Efficiently computing the step (24) for $k\in\mathcal{I}$ involves first computing $\mathbf{v}_k = 2\hat{\mathbf{b}}_k^{(t-1)} - \hat{\mathbf{c}}_k^{(t-1)}$ and then computing $\hat{\mathbf{a}}_k^{(t)} = \mathbf{v}_k - \mathbf{Q}_k\widehat{\mathbf{H}}_k\mathbf{v}_k + \mathbf{d}_k$.

[5]For the single-carrier case (i.e., when $N = 1$ and $S = 1$), SQUID-OFDM reduces to single-carrier SQUID [11, Sec. IV-B] and the per-iteration complexity reduces to $8BU + 4B$ real-valued multiplications (no split-radix FFT algorithm has to be computed for this case).



(a) 1-phase-bit SQUID-OFDM. (b) 2-phase-bit SQUID-OFDM. (c) 3-phase-bit SQUID-OFDM.

Fig. 2. Per-antenna SQUID-OFDM output before and after quantization for 16-QAM, $T = 20$, *SNR* $= 10$ dB, $B = 128$, $U = 16$, $S = 1200$, and $N = 4096$. The cross-markers correspond to the output before quantization, the circles to the quantized output, and the lines to the decision regions for the quantizer.

preprocessing complexity and the per-iteration complexity, we obtain the complexity for T iterations reported in Table I. Note that the computational complexity of both SQUID-OFDM and WF precoding scales *linearly* in the number of BS antennas B.

IV. Numerical Results

A. Simulation Parameters

Due to space constraints, we focus on a selected set of system parameters.[6] Specifically, we consider the case in which the BS is equipped with $B = 128$ antennas and serves $U = 16$ UEs. We consider long-term evolution (LTE)-inspired OFDM parameters [24] with $S = 1200$ occupied subcarriers and where $N = 4096$ (the oversampling ratio is $N/S = 4096/1200 \approx 3.41$). The subcarrier spacing is $\Delta f = 15$ kHz and the sampling rate is $f_s = N\Delta f = 61.44$ MHz. The set of occupied subcarriers is $\mathcal{I} = \{1, 2, \ldots, 600, 3497, 3498, \ldots, 4096\}$ and the set of guard subcarriers is $\mathcal{G} = \{0, 1, \ldots, 4096\} \setminus \mathcal{I}$. The entries of $\{\mathbf{H}_\ell\}_{\ell=0}^{L-1}$ are i.i.d. $\mathcal{CN}(0, 1/L)$ (Rayleigh fading). The number of taps is $L = 4$. We furthermore assume that the uth UE ($u = 1, 2, \ldots, U$) scales the received signal for each OFDM symbol by [12]

$$\beta_u = \frac{1}{\sqrt{\frac{1}{S}\sum_{k\in\mathcal{I}}|\hat{y}_k|^2 - N_0}}, \tag{27}$$

to obtain an estimate $\tilde{s}_{u,k} = \beta_u \hat{y}_{u,k}$ of $s_{u,k}$ for $k \in \mathcal{I}$.

B. Convergence and Complexity

We start by investigating the convergence of SQUID-OFDM for the 2-phase-bit case. Fig. 3 shows the complementary cumulative distribution function (CCDF) of the error-vector magnitude (EVM), with 16-QAM signaling for *SNR* $= 10$ dB. Here, the EVM for the uth UE ($u = 1, \ldots, U$) is defined as

$$EVM_u = \sqrt{\frac{\sum_{k\in\mathcal{I}}\left|s_{u,k} - \beta_u \hat{\mathbf{h}}_{u,k}^T \hat{\mathbf{x}}_k\right|^2}{\sum_{k\in\mathcal{I}}|s_{u,k}|^2}}, \tag{28}$$

where $\hat{\mathbf{h}}_{u,k}^T$ is the uth row of $\widehat{\mathbf{H}}$ and where β_u is given by (27). For reference, we also show the CCDF of the EVM with WF precoding for the 2-phase-bit case and for the *infinite-resolution* case (i.e., when $\mathbf{X}^{\text{WF}} = \widehat{\mathbf{Z}}^{\text{WF}}\mathbf{F}_N^H \in \mathbb{C}^{B\times N}$), respectively. Interestingly, we see that SQUID-OFDM with only one iteration already significantly outperforms WF precoding in terms of EVM. Furthermore, we see from Table I that for $T = 1$, the complexity of SQUID-OFDM is just about 3 times higher than that of WF precoding. We also see from Fig. 3 that by increasing the number of iterations from 20 to 100, SQUID-OFDM attains only marginal EVM gains. In what follows, we set the number of iterations to $T = 20$. In this case, SQUID-OFDM requires approximately 14 times more real-valued multiplications than the WF precoder, assuming the parameters given in Section IV-A.

[6]Our simulation framework is available for download from GitHub (https://github.com/quantizedmassivemimo/1bit_precoding_ofdm).

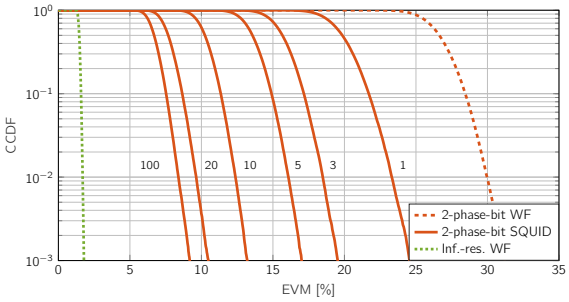
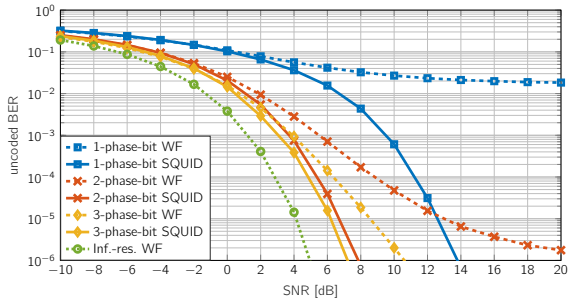


Fig. 3. CCDF of the EVM with 16-QAM; *SNR* $= 10$ dB, $B = 128$, $U = 16$, $S = 1200$, and $N = 4096$. The number next to the CCDF curves corresponds to the number of iterations for SQUID-OFDM.

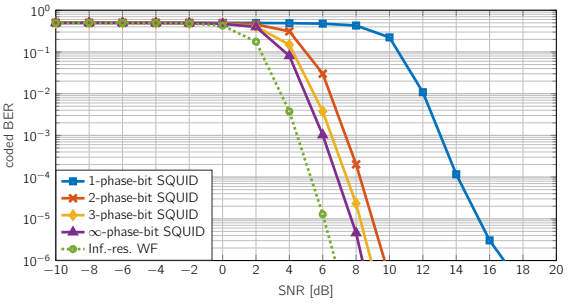
C. Error-Rate Performance

1) Uncoded BER: Fig. 4a shows the uncoded BER with 4-QAM for p-phase-bit ($p \in \{1, 2, 3\}$) SQUID-OFDM and WF precoding as a function of the SNR. We also show the uncoded BER with infinite-resolution WF precoding. We assume that the UEs perform symbol-wise nearest-neighbor decoding (i.e., each UE maps the received signal to the nearest constellation point in $\mathcal{O}$). We note that SQUID-OFDM outperforms WF precoding for all considered values of SNR and irrespectively of the number of phase bits. Interestingly, low uncoded BERs are supported even by 1-phase-bit SQUID-OFDM.

2) Coded BER: Fig. 4b shows the coded BER with 16-QAM for p-phase-bit ($p \in \{1, 2, 3, \infty\}$) SQUID-OFDM as a function of the SNR. For ∞-phase-bit SQUID-OFDM, the output after the last iteration is mapped to the set $\mathcal{X}_\infty^{B\times N}$ (no quantization).



(a) Uncoded BER with 4-QAM. SQUID-OFDM outperforms WF precoding irrespectively of the SNR and the number of phase bits.



(b) Coded BER with 16-QAM (rate-5/6 convolutional code). Low coded BERs are supported with SQUID-OFDM.

Fig. 4. Uncoded/coded BER as a function of SNR and the number of phase bits; $B = 128$, $U = 16$, $S = 1200$, and $N = 4096$. SQUID-OFDM significantly outperforms linear precoders and approaches infinite resolution performance.

At the BS, the information bits are encoded using a weak rate-5/6 convolutional code. Each codeword is randomly interleaved over 4800 bits (i.e., over the $S = 1200$ occupied subcarriers in an OFDM symbol). To detect the information bits, each UE performs soft-input max-log BCJR decoding. We note that low coded BERs are supported with SQUID-OFDM. Also, we note that 2-phase-bit SQUID-OFDM already offers performance close to that of ∞-phase-bit SQUID-OFDM.

V. Conclusions

We have proposed a nonlinear phase-quantized precoder called SQUID-OFDM for the massive MU-MIMO-OFDM downlink. The precoder extends the SQUID precoder in [11] to support OFDM, oversampling DACs, and arbitrary constant-envelope alphabets. SQUID-OFDM is shown to offer superior error-rate performance to linear precoders such as WF precoding at an increased computational complexity (three times or higher depending on the number of algorithm iterations). The constant-envelope transmit signals generated by SQUID-OFDM enable energy-efficient PAs, which is in stark contrast to the infinite-precision WF precoder whose PAR approximately ranges between 10 dB and 12 dB. Furthermore, for 2-phase-bit SQUID-OFDM, the amount of raw data that has to be fed to the DACs is 15.7 Gbit/s. In contrast, a traditional system that uses high-resolution DACs, e.g., 12-bit, must sustain raw baseband data rates that exceed 188 Gbit/s for the parameters considered in this paper, which seems impractical.

References

[1] F. Boccardi, R. W. Heath Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[2] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[3] C. Mollén, E. G. Larsson, and T. Eriksson, "Waveforms for the massive MIMO downlink: Amplifier efficiency, distortion, and performance," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5050–5063, Dec. 2016.

[4] S. S. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Commun.*, vol. 12, no. 2, pp. 56–65, Apr. 2005.

[5] S. K. Mohammed and E. G. Larsson, "Constant-envelope multi-user precoding for frequency-selective massive MIMO systems," *IEEE Commun. Lett.*, vol. 2, no. 5, pp. 547–550, Oct. 2013.

[6] A. Noll, H. Jedda, and J. A. Nossek, "PSK precoding in multi-user MISO systems," in *Proc. Int. ITG Workshop on Smart Antennas (WSA)*, Berlin, Germany, Mar. 2017, pp. 57–63.

[7] S. Jacobsson, G. Durisi, M. Coldrey, and C. Studer, "Linear precoding with low-resolution DACs for massive MU-MIMO-OFDM downlink," Sep. 2017. [Online]. Available: https://arxiv.org/abs/1709.04846

[8] ——, "Massive MU-MIMO-OFDM downlink with one-bit DACs and linear precoding," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Singapore, Dec. 2017.

[9] H. Jedda, J. A. Nossek, and A. Mezghani, "Minimum BER precoding in 1-bit massive MIMO systems," in *IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*, Rio de Janeiro, Brazil, Jul. 2016.

[10] A. L. Swindlehurst, A. K. Saxena, A. Mezghani, and I. Fijalkow, "Minimum probability-of-error perturbation precoding for the one-bit massive MIMO downlink," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 6483–6487.

[11] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized precoding for massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, Nov. 2017.

[12] ——, "Nonlinear 1-bit precoding for massive MU-MIMO with higher-order modulation," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 763–767.

[13] O. Castañeda, S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "1-bit massive MU-MIMO precoding in VLSI," *IEEE J. Emerging Sel. Topics Circuits Syst.*, vol. 7, no. 4, pp. 508–522, Dec. 2017.

[14] A. Li, C. Masouros, F. Liu, and A. L. Swindlehurst, "Massive MIMO 1-bit DAC transmission: A low-complexity symbol scaling approach," Nov. 2017. [Online]. Available: https://arxiv.org/abs/1709.08278

[15] L. Landau and R. C. de Lamare, "Branch-and-bound precoding for multiuser MIMO systems with 1-bit quantization," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 770–773, Dec. 2017.

[16] A. Nedelcu, F. Steiner, M. Staudacher, G. Kramer, W. Zirwas, R. Sisava Ganesan, P. Baracca, and S. Wesemann, "Quantized precoding for multi-antenna downlink channels with MAGIQ," in *Int. ITG Workshop on Smart Antennas (WSA)*, Bochum, Germany, Mar. 2017.

[17] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[18] P. L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, no. 6, pp. 964–979, Dec. 1979.

[19] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.

[20] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programming*, vol. 55, pp. 293–318, Apr. 1992.

[21] M. Wu, B. Yin, K. Li, C. Dick, J. R. Cavallaro, and C. Studer, "Implicit vs. explicit approximate matrix inversion for wideband massive MU-MIMO data detection," *Journal of Signal Processing Systems*, Dec. 2017.

[22] P. Duhamel and M. Vetterli, "Fast Fourier transforms: A tutorial review and a state of the art," *Signal Processing*, vol. 19, pp. 259–299, Apr. 1990.

[23] K. Li, R. R. Sharan, Y. Chen, T. Goldstein, J. R. Cavallaro, and C. Studer, "Decentralized baseband processing for massive MU-MIMO systems," *IEEE J. Emerging Sel. Topics Circuits Syst.*, vol. 7, no. 4, pp. 491–507, Dec. 2017.

[24] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE/LTE-Advanced for Mobile Broadband*. Academic Press, Dec. 2013.